Grenze International Journal of Engineering and Technology, June Issue

Predictive Modelling for Liver Disease Diagnosis using Machine Learning

Vidhya S G[1], Bhavana M R[2], Darshan Gowda S R[3], Prajwal K J[4] and Prateek J[5]
[1]Asst Prof, Dept. of Information Science and Engineering, ACU, BGSIT, India
Email: vidhyasg@bgsit.ac.in
[2-5]UG Student, Dept. of Information Science and Engineering, ACU, BGSIT, India

***Abstract*— Liver illness, affecting the human body primarily in the liver, is one of the most severe diseases. The liver plays a crucial role in waste removal, vitamin storage, and digestion. Early detection is vital in mitigating the risks associated with this lethal disease and potentially saving lives. Approximately 3.5 percent of the global population is affected by liver illness. Advancements in disease detection, particularly using machine learning classification techniques, have shown promise in improving outcomes. Other machine learning methods like artificial neural networks and convolutional neural networks are also being employed to address this challenge. These methods aim to increase the life expectancy of affected patients and prevent the development of chronic liver disease (CLD). The widespread use of barcodes, automation in commercial and government transactions, and advancements in data gathering systems have led to the accumulation of vast amounts of data. In response, proposed models utilize ensemble methods such as random forest, ADA boost, and gradient boost to achieve higher accuracy in disease detection. By combining these techniques, the model aims to enhance diagnostic capabilities and improve patient outcomes.**

I. INTRODUCTION

The liver, one of the largest organs in the human body, primarily functions to process food and eliminate harmful substances. Liver damage caused by viruses and alcohol consumption poses a significant threat to life. Various diseases such as hepatitis, cirrhosis, liver tumours, and liver cancer affect the organ. Fatty liver disease, characterized by the accumulation of fat in the liver, is prevalent in India, with over 10 million cases reported annually. Diagnosis often requires testing due to the lack of noticeable symptoms. Liver disease encompasses any condition affecting the liver's function, which can have serious repercussions on overall health. The liver performs numerous critical functions in the body, and its dysfunction can lead to severe consequences. Artificial Intelligence (AI), a form of computerized reasoning enabled by the ability of computer programs to learn and apply knowledge, is increasingly utilized across various domains. AI, which includes components such as machine learning and deep learning, is applied extensively, aiming to alleviate clinicians' workload and improve medical outcomes. Major causes of liver disease include obesity, exposure to toxic gases, consumption of contaminated food, excessive medication or food intake, and alcohol abuse. The research aims to utilize machine learning techniques for the prediction and detection of liver disorders, offering several anticipated outcomes:

1. Prediction of liver disease occurrence.
2. Early detection of liver disease to prevent further complications.
3. Reduction in liver disease-related mortality rates.

Grenze ID: 01.GIJET.10.2.743

4. Enhanced accuracy in the diagnosis of liver disease.
These outcomes represent significant advancements in liver disease management, potentially leading to improved patient outcomes and reduced burden on healthcare systems.

II. Motivation

The paper focuses on predicting liver disease using machine learning, addressing a crucial and challenging issue in the medical field and society at large. Despite existing research in this area, there are gaps and limitations that need to be addressed. The motivation behind the paper is to propose an effective solution to this problem based on specific requirements. The main contributions of the paper include fast prediction, high accuracy, and cost reduction, showcasing the advantages of the proposed approach over existing state-of-the-art methods. This paper is expected to have significant implications for both theory and practice in the fields of machine learning, medicine, and society. By offering a solution that improves prediction speed, accuracy, and cost-effectiveness, this paper has the potential to enhance medical diagnostics and patient care. Additionally, the findings may inform future research and contribute to advancements in machine learning methodologies for healthcare applications. Overall, the paper holds promise for positively impacting both medical practice and broader societal outcomes.

III. Literature Survey

Several papers have been proposed on the prediction of liver disease using machine learning techniques. Let's summarize the key points from each paper:
1. R.T. Umbare et al. compared four algorithms (Logistic Regression, K-Nearest Neighbors, SVM, and Artificial Neural Network) for liver disease prediction. They found that SVM achieved the highest accuracy of 75.04%.
2. Sai Rohith Tanuku et al. introduced a novel ensemble technique combining Random Forest, XGBoost, and Gradient Boost algorithms. They claimed higher accuracy than existing methods but raised concerns about potential overfitting.
3. Rakshith D B et al. utilized non-invasive and readily available data for liver disease prediction, evaluating six algorithms. They reported Decision Stump as the most accurate with 91.7% accuracy.
4. Deepika Bhupathi et al. proposed a method using text mining to understand traditional Chinese medicine pathogenesis of non-alcoholic fatty liver disease. They employed an autoencoder network and a classifier, focusing more on mining than machine learning, which could lead to lower accuracy and erroneous predictions.
5. A.K.M S R F. M. J M Shamrat et al. aimed to reduce the cost of chronic liver disease diagnosis by evaluating the performance of various machine learning algorithms. They found varying accuracies ranging from 53% to 75% for different algorithms.
6. Dr. S Vijayarani et al. focused on predicting liver diseases using Naïve Bayes and SVM algorithms. While SVM showed promising results, the model was prone to overfitting and required a large amount of data for training.
Overall, these papers contribute valuable insights into liver disease prediction using machine learning techniques, showcasing various approaches and algorithms with differing levels of accuracy and complexity.

IV. Proposed System

The proposed model implements the prediction of liver disease using Ensemble Technique, aiming to alleviate the burden on doctors by analysing patient conditions using machine learning techniques for quicker responses. Ensemble techniques combine multiple models to achieve better results, often outperforming individual models. Specifically, our system utilizes the Gradient Boosting Classifier + AdaBoost Classifier (Ensemble Technique) to predict the presence of liver disease symptoms based on given datasets of individuals. The primary goal is to accurately predict the disease status using Ensemble Technique, which would greatly benefit hospitals, professional doctors, and general users in detecting liver diseases efficiently. This model has the potential to revolutionize disease detection and could save lives by facilitating early diagnosis and treatment.
The entire work is dedicated to developing a predictive model using given datasets to aid in the prevention and treatment of liver diseases. The datasets follow a specific path, as indicated in the following diagram:
The diagram illustrating the path of datasets, depicting the flow of data through the system, including data preprocessing, model training, and prediction generation.

V. System Design

This outlines the step-by-step process for developing a machine learning model for liver disease prediction:

1. Collect Data: Gathering data is the initial step, critical for the model's performance. Techniques like web scraping or manual collection from hospitals or medical centres are employed to obtain data relevant to liver disease prediction.
2. Dataset: The dataset comprises 590 individual data points with 11 columns containing necessary information for liver disease prediction. Only selected data from medical test reports are included.
3. Pre-Processing and Feature Selection: The dataset undergoes thorough pre-processing to remove irrelevant, incomplete, or inconsistent information. Python libraries are utilized for data pre-processing, including transforming string data into numerical values to fit data with classifiers. Feature selection is also carried out to reduce the data size, optimizing the dataset for analysis.
4. Model Selection: The Gradient Boosting Classifier + AdaBoost Classifier (Ensemble Technique) using machine learning algorithms is chosen for its superior accuracy. Achieving a test set accuracy of 92.1%, this algorithm is implemented for liver disease prediction.
5. Ensemble Technique: Ensemble methods are employed to enhance model accuracy by combining multiple models rather than relying on a single model. This approach significantly boosts the accuracy of the prediction results, thus increasing the effectiveness of the machine learning technique.
6. Performance Analysis: Various metrics are utilized to evaluate the classification model's performance. Accuracy, representing the ratio of correctly predicted samples to the total samples, and classification error, indicating the error present in the samples, are among the metrics considered for performance analysis.
Overall, this systematic approach ensures the development of an accurate and reliable machine learning model for liver disease prediction, contributing to improved diagnosis and patient care.

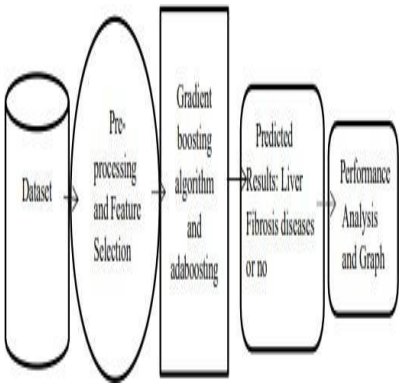

Fig 1. System Architecture

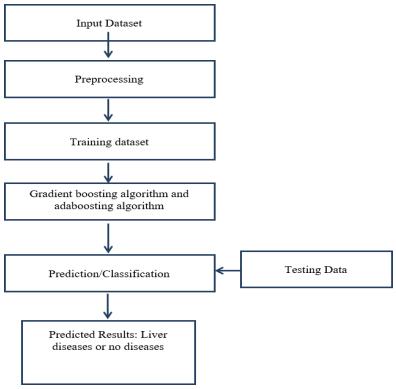

Fig.2 Data flow Diagram

VI. ALGORITHMS

[1] In our approach, we utilize the Gradient Boosting Regressor for regression problems, while for classification tasks, we opt for the Gradient Boosting Classifier. The primary distinction lies in the choice of loss function. The fundamental objective is to minimize the loss function by iteratively adding weak learners using gradient descent. For regression tasks, various loss functions like Mean Squared Error (MSE) are employed, whereas for classification, different loss functions such as log-likelihood are utilized.
[2] Adaptive Boosting, commonly known as AdaBoost, is an Ensemble Method in Machine Learning. It is often paired with decision trees having only one level, which are referred to as Decision Stumps. These decision stumps serve as the base learners in AdaBoost, contributing to its effectiveness in creating a strong ensemble model.

VII. CONCLUSION

After completing the entire process, from data preparation to preprocessing to selecting and combining different models, the obtained output holds significant potential for the medical industry, particularly in predicting liver disease accurately. This proposed model has demonstrated impressive performance, achieving a Train Accuracy of 100% and Test Accuracy of 92% and above. Such high accuracy is invaluable in the medical field, as it can potentially save lives by aiding in early detection and treatment of liver disease. The extensive work conducted can greatly benefit the entire medical industry. The accuracy achieved can be leveraged in real-time web applications developed by professionals, enabling timely and accurate predictions for liver disease. Additionally, further research can explore the use of different ensemble methods and conduct comparative analyses of various

machine learning techniques and ensemble techniques. This would provide deeper insights into the effectiveness of different approaches and help optimize prediction models even further. Ultimately, the culmination of these efforts holds immense promise for improving patient care and outcomes in the medical field.
In conclusion, the proposed model holds promise in revolutionizing disease prediction and could significantly contribute to improving healthcare outcomes for patients with liver diseases.

REFERENCES

[1] M. K. T. G. K. M. K. B.Muruganantham, "Liver Disease Prediction Using Classification Algorithms".
[2] S. E. A. I. E.-D. Engy El-Shafeiy," Prediction of Liver Diseases Based on Machine Learning Technique for Big Data".
[3] Z. T. Joy Roy," A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning".
[4] S. R. D. K. Vasan Durai," Liver disease prediction using Machine Learning".
[5] M. S. A. K. G. S. P. 1Rakshith D B," Liver Disease Prediction System using Machine Learning Techniques," 2021.
[6] J. C. M. J. M. E. I. Joel Jacob," Diagnosis of Liver Disease Using Machine Learning Techniques," 2018.
[7] L. J. T. Banu Priyal," Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms".
[8] R. T. Ribeiro, R. T. Marinho and J. M. Sanches," Classification and Staging of Chronic Liver Disease from Multimodal Data," in IEEE Transactions on Biomedical Engineering, vol. 60, no. 5, pp. 1336-1344, May 2013, doi: 10.1109/TBME.2012.2235438.
[9] J. Lara, Y. Khudyakov, F. Xavier L o´pezLabrador, F. Gonzalez Candelas and M. Berenguer," Hepatitis C virus genetic association to rate of liver fibrosis progression," 2013 IEEE 3rd International Conference on Computational Advances in Bio and medical Sciences (ICCABS), 2013, pp. 1-1, doi:10.1109/ICCABS.2013.6629225
[10] S. Ambesange, V. A, R. Uppin, S. Patil and V. Patil," Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques," 2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2020, pp. 98-102, doi: 10.1109/CCEM50674.2020.00030
[11] F. Destrempes et al.," Machine learning based on quantitative ultrasound for assessment of chronic liver disease," 2020 IEEE International Ultrasonics Symposium (IUS),2020, pp.1- 3, doi:10.1109/IUS467 67.2020.9251512.
[12] H. -q. Zhang et al.," Using text mining to understand traditional Chinese medicine pathogenesis of nonalcoholic fatty liver disease," 2013 IEEE International Conference on Bioinformatics and Biomedicine, 2013, pp. 311- 314, doi: 10.1109/BIBM.2013.6732702.
[13] Shah and S. Modi,"Comparative Analysis of Psychometric Prediction System," 2021 Smart Technologies, Communication and Robotics (STCR),2021,pp.1-5,doi:10.1109/STCR51658.202 1 .9588950.
[14] Sivasangari, B. J. Krishna Reddy, A. Kiran and P. Ajitha,"Diagnosis of Liver Disease using Machine Learning Models," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 627-630,doi:10.1109/ISMAC49090.2020.9243375.